Provable Security of Block Ciphers Against Linear Cryptanalysis - a Mission Impossible?

An Experimental Review of the Practical Security Approach and the Key Equivalence Hypothesis in Linear Cryptanalysis

Gilles Piret[1], François-Xavier Standaert[2]*

[1] Oberthur Card Systems, Nanterre, France
[2]UCL Crypto Group, Microelectronics Laboratory, Louvain-la-Neuve, Belgium

Abstract. In this paper, we are concerned with the security of block ciphers against linear cryptanalysis and discuss the distance between the so-called practical security approach and the actual theoretical security provided by a given cipher. For this purpose, we present a number of illustrative experiments performed against small (*i.e.* computationally tractable) ciphers. We compare the linear probability of the best linear characteristic and the actual best linear probability (averaged over all keys). We also test the key equivalence hypothesis. Our experiments illustrate both that provable security against linear cryptanalysis is not achieved by present design strategies and the relevance of the practical security approach. Finally, we discuss the (im)possibility to derive actual design criteria from the intuitions underlined in these experiments.

Keywords: Symmetric Cryptography, Block Ciphers, Linear Cryptanalysis.
AMS Classification: 94A60

1 Introduction

The linear cryptanalysis [15, 22] is one of the most powerful attacks against block ciphers. However, although a number of commonly accepted strategies have been developed to provide practical security against such adversaries (most famously, the wide-trail strategy [6] that has been used for the design of the AES Rijndael [7]), the foundations of these important techniques are mainly based on a number of practically acceptable but theoretically disputable hypotheses. In addition, actual solutions to counteract linear cryptanalysis are frequently based on heuristics rather than on a sound theoretical framework allowing provable security. A significant reason for these practice-oriented approaches is the difficulty of analyzing and understanding the statistical properties of linear approximations within actual block ciphers, which frequently results in computationally intensive tasks. Another reason is the need of flexible and efficient solutions for the practical instances of block ciphers targeted for real applications, which does not always fit with elegant theoretical constructions. As a consequence, the present state-of-the art in block cipher design is mainly based on a combination of engineering principles led by reasonable theoretical guidelines.

* Postdoctoral researcher of the Belgian Fund for Scientific Research (FNRS).

In this context, a good assessment of the hypotheses used for the evaluation of linear cryptanalysis as well as a measurement of the distance between actual constructions and theoretical expectations are of particular interest. For this purpose, this paper intends to discuss certain important tools in present block cipher design, in function of the traditional parameters : block cipher size and number of rounds. Because of the previously mentioned theoretical difficulties, we considered a number of experimental investigations, built upon small ciphers (*i.e.* we considered computationally tractable block sizes). Then, from these empirical observations, we aim to underline certain useful intuitions for the understanding (and possibly the design) of block ciphers.

First, we designed experiments to evaluate the relevance of the *use of characteristics for arguing the security of a construction*, *i.e.* we measured the distance between theoretical and practical security (as defined by Knudsen in [17]). We show experimental evidence that the theoretical security limits of a cipher essentially depend on its block size. We also put forward that the practical security approach does not relate to the existence of a linear attack with low data complexity but to the difficulty of finding it. Second, we consider the *key equivalence hypothesis* [13] and confirm that it is better fulfilled for large key sizes, as long as the block cipher is reasonably designed. These experiments highlight another aspect of the practical security approach: if the best linear approximation of a given cipher is key-dependent, it can hardly be exploited by an actual adversary (this fact is typically used in the decorrelation theory [33]). We finally illustrate that the theoretical security bound of a block cipher is obtained when its linear approximations are reaching a stationary area, in which adding a round to the cipher is equivalent to changing the key. We discuss the (im)possibility to derive practical *design criteria* for block ciphers from these observations.

Note that the authors of the paper do not claim the novelty of their conclusions and acknowledge the tutorial nature of this work. As a matter of fact, our experiments confirm a number of intuitive views that can be found in former papers. However, we believe that such an experimental approach is useful for the understanding of the complex mechanisms involved in linear cryptanalysis and raises questions about the design criteria to counteract such attacks.

2 Target Ciphers and Notations

Typical (key alternating) block ciphers are the Substitution Permutation Network (SPN) and the Feistel structure that are represented in Figure 1. According to the usual terminology in use for block ciphers, their block size is n and number of rounds is R. The SPN round is divided into a key addition (bitwise XOR) with a round key K_i, n_s substitution boxes (S-boxes) of size b and a linear diffusion layer. The same round structure is used in the Feistel cipher as a non-linear function F. In the following, our experiments will only consider SPNs, but similar investigations could be carried on with Feistel ciphers. The exact specifications of the S-boxes and diffusion layer used in our experiments are given in appendix A.

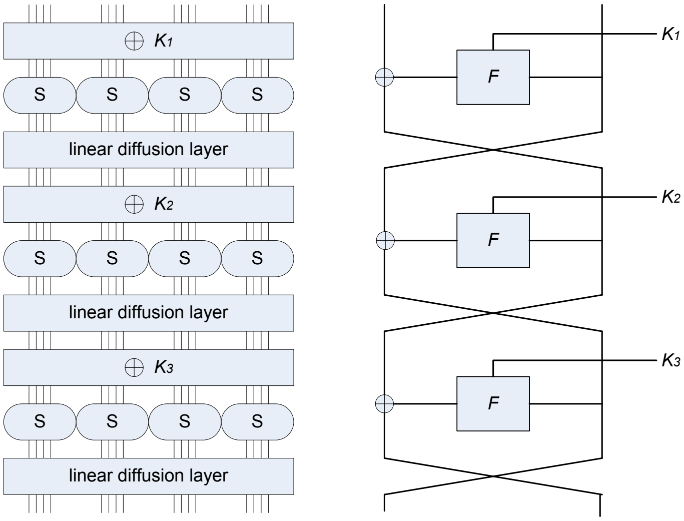

Fig. 1: Exemplary SPN and Feistel structures.

In general, the resistance of a cipher against linear cryptanalysis depends on its non-linearity and thus the one of its components (*e.g.* the S-boxes). Several definitions exist that capture the extent to which a Boolean mapping is non-linear. We will use the linear probability [32].

Definition 1. *Let $F : \{0,1\}^n \rightarrow \{0,1\}^n$ be a bijection and $\boldsymbol{a}$,$\boldsymbol{b}$ be two masks $\in \{0,1\}^n$. If $X \in \{0,1\}^n$ is a uniformly distributed random variable, then the linear probability $LP(\boldsymbol{a},\boldsymbol{b})$ is defined as*

$$LP(\boldsymbol{a},\boldsymbol{b}) = (2 \cdot \Pr_X\{\boldsymbol{a} \bullet X = \boldsymbol{b} \bullet F(X)\} - 1)^2, \tag{1}$$

where $\bullet$ denotes the usual scalar product. If F is parametrized by a key K, we write $LP(\boldsymbol{a},\boldsymbol{b};K)$ and the expected linear probability $ELP(\boldsymbol{a},\boldsymbol{b})$ is defined as:

$$ELP(\boldsymbol{a},\boldsymbol{b}) = \underset{K}{\boldsymbol{E}}\ [LP(\boldsymbol{a},\boldsymbol{b};K)] \tag{2}$$

3 Related Works

As a number of research papers on iterative block ciphers, our following results assume independent round keys. We did not include the verification of this assumption in our experiments since it has already been experimentally tested by Knudsen and Mathiassen in [19]. They showed that the key schedule actually matters, but the conclusions drawn when independent round keys are used should still reasonably hold with a good complex key schedule. Also, we only considered the Markov ciphers defined in [21]: a Markov cipher is a cipher for which the linear (and differential) probabilities of different rounds are independent of each other, assuming uniformly random keys.

In this section, we briefly summarize existing works related to the provable and practical security of block ciphers against linear cryptanalysis. Let $\tilde{K}$ denote the vector of subkeys $\{K_1, K_2, \ldots, K_R\}$ used in the R rounds of a target block cipher. Ideally, evaluating the security of the cipher against linear cryptanalysis would require to compute the value:

$$\max_{\mathbf{a},\mathbf{b}} \; LP(\mathbf{a}, \mathbf{b}; \tilde{K}) \tag{3}$$

It directly yields the approximated data complexity of the attack [22], namely:

$$N_L \approx \frac{c}{\max \; LP(\mathbf{a}, \mathbf{b}; \tilde{K})}, \tag{4}$$

where c is a small constant value. Unfortunately, the direct computation of (3) is generally infeasible, both for computational reasons and because of an unknown key. As a consequence, research works on linear cryptanalysis are usually based on two important assumptions. First, a common approximation to solve the key dependencies problem is to compute the expected value $ELP(\mathbf{a}, \mathbf{b})$ and to assume that for almost all values $\tilde{K}$, we have:

$$LP(\mathbf{a}, \mathbf{b}; \tilde{K}) \approx ELP(\mathbf{a}, \mathbf{b}) \tag{5}$$

Harpes *et al.* called this assumption the *key equivalence hypothesis* [13]. Using this hypothesis, a theory of provable security against linear cryptanalysis has been developed. It notably gave rise to the design principles used in the MISTY1 algorithm [23] or the CS cipher [32]. The main contribution of this approach is to provide bounds on the expected linear probabilities of a cipher. Its main limitation is the computational difficulty of finding tight bounds when the number of block cipher rounds increases. As a consequence, a more practical view of the security against linear cryptanalysis was developed in parallel, based on a second assumption using the concept of characteristic.

Definition 2. *A one-round characteristic for the round i is a pair of n-bit vectors $\langle \boldsymbol{a}^i, \boldsymbol{b}^i \rangle$ respectively corresponding to the input and output masks for this round. A R-round characteristic for rounds $1 \ldots R$ is a $(R+1)$-tuple of n-bit vectors $\Omega = \langle \boldsymbol{a}^1, \boldsymbol{a}^2, \ldots, \boldsymbol{a}^{R+1} \rangle$, where $\langle \boldsymbol{a}^i, \boldsymbol{a}^{i+1} \rangle$ respectively correspond to the input and output masks for the round i.*

Definition 3. *Given a vector of independent subkeys $\tilde{K}$, the linear characteristic probability and expected linear characteristic probability of a R-round characteristic Ω are defined as*[1]*:*

$$LCP(\Omega, \tilde{K}) = \prod_{i=1}^{R} LP(\boldsymbol{a}^i, \boldsymbol{a}^{i+1}, K_i) \tag{6}$$

$$ELCP(\Omega) = \prod_{i=1}^{R} ELP(\boldsymbol{a}^i, \boldsymbol{a}^{i+1}) \tag{7}$$

These definitions essentially state that for a Markov cipher and assuming independent round keys, the probability of a R-round characteristic can be computed as a product of 1-round characteristics probabilities. In order to ensure practical security, a designer typically runs an algorithm to search the characteristic $\tilde{\Omega}$ such that $ELCP(\tilde{\Omega})$ is maximal and then assumes:

$$ELP(\mathbf{a}, \mathbf{b}) \approx ELCP(\tilde{\Omega}) \tag{8}$$

Knudsen calls a block cipher practically secure if the data complexity determined by this method is prohibitive[2] [17]. Obviously, such an approximation is only valid to a certain extent and it may give rise to false intuitions. For example, increasing the number of rounds always reduces the linear characteristic probabilities while the actual expected linear probability of a cipher cannot be decreased below a certain threshold, depending on its block size. In order to avoid such an overestimation of the attack complexities, Nyberg consequently introduced the concept of linear hull [24] that is defined as follows.

Definition 4. *Given input and output masks* $\boldsymbol{a}$,$\boldsymbol{b}$*, the approximated linear hull* $ALH(\boldsymbol{a}, \boldsymbol{b})$ *is the set of all R-round characteristics having* $\boldsymbol{a}$ *as input mask for round 1 and* $\boldsymbol{b}$ *as output mask for round R.*

Evaluating the linear hull effect was then another direction to derive provably secure Markov ciphers against linear cryptanalysis, since :

$$ELP(\mathbf{a}, \mathbf{b}) = \sum_{\Omega \in ALH(\mathbf{a}, \mathbf{b})} ELCP(\Omega) \tag{9}$$

However, similarly to other theories of provable security, the estimation of the linear hulls hardly results on tight bounds when the number of ciphers rounds increases [16]. In addition, Equation (9) still assumes the key equivalence.

Based on this short state-of-the art, the following sections aim to experimentally and intuitively evaluate (1) the distance between practical and provable security and (2) the validity of the key equivalence hypothesis. Thereafter, we discuss the possibility to derive actual design criteria from empirical measurements of the linear characteristics within block ciphers.

Note that various other works can be related to this line of research. For example, Selçuk [31] tried to assess the pertinence of the practical security approach, with experiments similar to ours. However his work only deals with poor diffusion layers and characteristics with few active S-boxes in the context of Feistel networks. He also considers reduced versions of RC5, which is a totally different cipher: it has no S-boxes, non-linearity comes from data-dependent rotations. On the contrary, our experiments consider SPNs of various sizes with good and bad diffusion layers. Selçuk also shows that trying to evaluate the linear probability of a linear approximation by means of statistical sampling (*i.e.* by measuring the LP on a few plaintexts only) is not going to succeed.

The recent report of Daemen and Rijmen [8] investigates the statistical distributions of the fixed key linear probabilities $LP(\mathbf{a},\mathbf{b},\tilde{K})$ and notably show that they have a Gamma distribution over the keys, with mean $ELP(\mathbf{a},\mathbf{b})$. We similarly focus on the key equivalence hypothesis and most of our experimental results can be related to this framework. However, we put a stronger focus on the distributions of $\max_{\mathbf{a},\mathbf{b}} LP(\mathbf{a},\mathbf{b},\tilde{K})$ (*i.e.* we don't fix the input and output masks but select the worst cases). In addition, our aim is to evaluate the influence of the block cipher parameters (*i.e.* block size, number of rounds) onto these distributions. Finally, decorrelation theory [33] aims at preventing the use of the key equivalence hypothesis in an attack. As a matter of fact, a *decorrelation module* is a key-dependent transformation that makes the linear probability (and differential probability) of a given approximation highly key-dependent, so that any attack that chooses the input and output masks *a priori* will fail... However, it does not prevent all kinds of linear and differential attacks [20] [34].

4 Specification of the Cipher Components

We use bijective S-boxes of sizes ranging from $b = 4$ to 8, with linear probabilities:

b	4	6	8
LP	$(2 \cdot \frac{1}{4})^2$	$(2 \cdot \frac{3}{16})^2$	$(2 \cdot \frac{1}{8})^2$

We use two types of diffusion layers: (1) an optimized transform with maximum branch number [7] $n_s + 1$ denoted as M; (2) a simpler wire crossing layer balancing the S-box output bits, denoted as C. From these definitions, we denote an n-bit SPN using b-bit S-boxes and a diffusion layer D as $SPN^D_{n,b}$. In particular, we consider the following sizes: $SPN_{8,4}$, $SPN_{12,4}$, $SPN_{12,6}$, $SPN_{16,4}$, $SPN_{16,8}$.

5 Limitations of the Practical Security Approach

In order to evaluate the extent to which the practical security approach is meaningful for actual block ciphers, we first computed the following quantities:

$$\max_{char} := \max_{\Omega} \; ELCP(\Omega), \tag{10}$$

$$\max_{hull} := \mathbf{E}_{\tilde{K}} \; \max_{\mathbf{a},\mathbf{b}} \; LP(\mathbf{a},\mathbf{b};\tilde{K}), \tag{11}$$

for various SPNs. The results of these computations are summarized in Tables 1, 2 in which we can observe the following facts:

1. After a sufficient number of rounds, the average best approximation of a given cipher (*i.e.* $\mathbf{E}_{\tilde{K}} \max LP$) only depends on its block size n (as long as the cipher does not contain obvious weaknesses, *e.g.* does not contain any linear approximation with probability one). Illustratively, our experiments suggest that the average best linear probability of a 16-bit (*resp.* 12-bit) cipher is $6.30 \cdot 10^{-4}$ (*resp.* $7.44 \cdot 10^{-3}$).

2. By contrast (and under the same conditions of "good cipher") the probability of the best characteristic goes on decreasing with the number of rounds R.
3. The number of rounds necessary for a given cipher to reach its minimum $\mathbf{E}_{\tilde{K}} \max LP$ value depend on its S-boxes and diffusion properties. For example, this limit value for $\mathbf{E}_{\tilde{K}} \max LP$ is faster achieved with 8-bit S-boxes (*resp.* a good diffusion layer) than 4-bit ones (*resp.* a bad one).
4. We note that in some cases, the expected best linear probability becomes greater when one more round is added. This effect is due to the important linear hull effect, and is observable only for a small number of rounds.

Table 1: Comparison between the best expected linear characteristic probability and the expected best linear probability for various SPNs with a bad diffusion layer.

	2×4		3×4		2×6		4×4		2×8	
# rounds	$\max_{char}$	$\max_{hull}$	$\max_{char}$	$\max_{hull}$	$\max_{char}$	$\max_{hull}$	$\max_{char}$	$\max_{hull}$	$\max_{char}$	$\max_{hull}$
1	$2.50 \cdot 10^{-1}$	$2.50 \cdot 10^{-1}$	$2.50 \cdot 10^{-1}$	$2.50 \cdot 10^{-1}$	$1.40 \cdot 10^{-1}$	$1.40 \cdot 10^{-1}$	$2.50 \cdot 10^{-1}$	$2.50 \cdot 10^{-2}$	$6.25 \cdot 10^{-2}$	$6.25 \cdot 10^{-3}$
2	$6.25 \cdot 10^{-2}$	$1.95 \cdot 10^{-1}$	$6.25 \cdot 10^{-2}$	$1.94 \cdot 10^{-1}$	$1.97 \cdot 10^{-2}$	$3.98 \cdot 10^{-2}$	$6.25 \cdot 10^{-2}$	$6.25 \cdot 10^{-2}$	$2.26 \cdot 10^{-3}$	$7.15 \cdot 10^{-4}$
3	$1.56 \cdot 10^{-2}$	$1.09 \cdot 10^{-1}$	$1.56 \cdot 10^{-2}$	$5.79 \cdot 10^{-2}$	$1.34 \cdot 10^{-3}$	$1.42 \cdot 10^{-2}$	$1.56 \cdot 10^{-2}$	$7.09 \cdot 10^{-2}$	$7.72 \cdot 10^{-5}$	$9.54 \cdot 10^{-4}$
4	$3.90 \cdot 10^{-3}$	$7.93 \cdot 10^{-2}$	$3.90 \cdot 10^{-3}$	$3.04 \cdot 10^{-2}$	$9.09 \cdot 10^{-5}$	$7.48 \cdot 10^{-3}$	$3.90 \cdot 10^{-3}$	$3.05 \cdot 10^{-2}$	$2.10 \cdot 10^{-6}$	$6.29 \cdot 10^{-4}$
5	$9.76 \cdot 10^{-4}$	$7.66 \cdot 10^{-2}$	$9.76 \cdot 10^{-4}$	$1.47 \cdot 10^{-2}$	$8.18 \cdot 10^{-6}$	$7.45 \cdot 10^{-3}$	$9.76 \cdot 10^{-4}$	$1.28 \cdot 10^{-2}$	$6.30 \cdot 10^{-8}$	$6.30 \cdot 10^{-4}$
6	$2.44 \cdot 10^{-4}$	$7.64 \cdot 10^{-2}$	$2.44 \cdot 10^{-4}$	$8.50 \cdot 10^{-3}$	$7.99 \cdot 10^{-7}$	$7.42 \cdot 10^{-3}$	$2.44 \cdot 10^{-4}$	$5.10 \cdot 10^{-3}$	$7.54 \cdot 10^{-10}$	$6.31 \cdot 10^{-4}$
7	$6.10 \cdot 10^{-5}$	$7.64 \cdot 10^{-2}$	$6.10 \cdot 10^{-5}$	$7.48 \cdot 10^{-3}$	$7.80 \cdot 10^{-8}$	$7.44 \cdot 10^{-3}$	$6.10 \cdot 10^{-5}$	$2.38 \cdot 10^{-3}$	$1.94 \cdot 10^{-11}$	$6.30 \cdot 10^{-4}$
8	$1.52 \cdot 10^{-5}$	$7.64 \cdot 10^{-2}$	$1.52 \cdot 10^{-5}$	$7.44 \cdot 10^{-3}$	$3.95 \cdot 10^{-9}$	$7.43 \cdot 10^{-3}$	$1.52 \cdot 10^{-5}$	$1.13 \cdot 10^{-3}$	$4.60 \cdot 10^{-13}$	$6.30 \cdot 10^{-4}$

Intuitively, these experiments suggest that the linear hull effect increases with the number of rounds R and is finally dominant as soon as the cipher is practically secure (according to Knudsen's definition). Otherwise said, for practically secure ciphers, the security bound (*i.e.* data complexity for a successful attack) approximated with the best linear characteristic is no longer meaningful. These experiments also underline that in theory and for a given cipher and key (*i.e.* considering the cipher as an n-bit S-box), a low data complexity linear cryptanalysis is always possible (*i.e.* a data complexity lower than 2^n). For example, our 16-bit ciphers include (on average) an approximation with linear probability $6.30 \cdot 10^{-4}$ that would give rise to an attack of approximated data complexity $2^{10.6}$. However, this does *not* mean that the practical security approach is not good for designing ciphers. As a matter of fact, the practical security approach is not aimed to prevent the existence of a linear attack, but makes it difficult to actually find and exploit the best linear approximations, for computational reasons. This is because, for practically secure ciphers, they cannot be found by chaining small approximations anymore. And searching them exhaustively in a cipher for large n values has a complexity in $\mathrm{O}(2^{3n})$. Additionally, as far as Matsui's second algorithm is concerned [22] and for practically secure ciphers, it is unlikely that such a best approximation can give rise to a guess on a few key bits and therefore to a practical attack with low time complexity.

Note that the third observation typically relates to the quality of the key scheduling algorithm too. That is, ciphers with well-designed, complex key schedules will reach the minimum $\mathbf{E}_{\tilde{K}} \max LP$ value faster than ciphers with poorly designed key schedules. This was experimentally confirmed in the previously mentioned paper of Knudsen and Mathiassen [19].

Table 2: Comparison between the best expected linear characteristic probability and the expected best linear probability for various SPNs with a good diffusion layer.

	2×4		3×4		2×6		4×4		2×8	
# rounds	$\max_{char}$	$\max_{hull}$	$\max_{char}$	$\max_{hull}$	$\max_{char}$	$\max_{hull}$	$\max_{char}$	$\max_{hull}$	$\max_{char}$	$\max_{hull}$
1	$2.50 \cdot 10^{-1}$	$2.50 \cdot 10^{-1}$	$2.50 \cdot 10^{-1}$	$2.50 \cdot 10^{-1}$	$1.40 \cdot 10^{-1}$	$1.40 \cdot 10^{-1}$	$2.50 \cdot 10^{-1}$	$2.50 \cdot 10^{-3}$	$6.25 \cdot 10^{-2}$	$6.25 \cdot 10^{-4}$
2	$1.56 \cdot 10^{-2}$	$6.88 \cdot 10^{-2}$	$3.90 \cdot 10^{-3}$	$1.89 \cdot 10^{-2}$	$2.78 \cdot 10^{-3}$	$8.84 \cdot 10^{-3}$	$9.76 \cdot 10^{-4}$	$3.80 \cdot 10^{-4}$	$1.86 \cdot 10^{-4}$	$8.05 \cdot 10^{-4}$
3	$3.90 \cdot 10^{-3}$	$7.62 \cdot 10^{-2}$	$9.76 \cdot 10^{-4}$	$7.48 \cdot 10^{-3}$	$2.71 \cdot 10^{-4}$	$7.44 \cdot 10^{-3}$	$2.44 \cdot 10^{-4}$	$6.43 \cdot 10^{-4}$	$3.23 \cdot 10^{-6}$	$6.31 \cdot 10^{-4}$
4	$2.44 \cdot 10^{-4}$	$7.63 \cdot 10^{-2}$	$1.52 \cdot 10^{-5}$	$7.45 \cdot 10^{-3}$	$3.72 \cdot 10^{-6}$	$7.43 \cdot 10^{-3}$	$9.53 \cdot 10^{-7}$	$6.32 \cdot 10^{-4}$	$1.00 \cdot 10^{-8}$	$6.32 \cdot 10^{-4}$
5	$6.10 \cdot 10^{-5}$	$7.64 \cdot 10^{-2}$	$3.81 \cdot 10^{-6}$	$7.44 \cdot 10^{-3}$	$1.49 \cdot 10^{-7}$	$7.44 \cdot 10^{-3}$	$5.96 \cdot 10^{-8}$	$6.31 \cdot 10^{-4}$	$8.71 \cdot 10^{-11}$	$6.29 \cdot 10^{-4}$
6	$3.81 \cdot 10^{-6}$	$7.64 \cdot 10^{-2}$	$5.96 \cdot 10^{-8}$	$7.43 \cdot 10^{-3}$	$3.47 \cdot 10^{-9}$	$7.44 \cdot 10^{-3}$	$9.31 \cdot 10^{-10}$	$6.31 \cdot 10^{-4}$	$3.55 \cdot 10^{-13}$	$6.30 \cdot 10^{-4}$
7	$9.53 \cdot 10^{-7}$	$7.64 \cdot 10^{-2}$	$3.72 \cdot 10^{-9}$	$7.43 \cdot 10^{-3}$	$1.28 \cdot 10^{-10}$	$7.44 \cdot 10^{-3}$	$5.82 \cdot 10^{-11}$	$6.30 \cdot 10^{-4}$	$2.71 \cdot 10^{-15}$	$6.30 \cdot 10^{-4}$
8	$5.96 \cdot 10^{-8}$	$7.64 \cdot 10^{-2}$	$2.32 \cdot 10^{-10}$	$7.44 \cdot 10^{-3}$	$2.07 \cdot 10^{-12}$	$7.43 \cdot 10^{-3}$	$9.09 \cdot 10^{-13}$	$6.32 \cdot 10^{-4}$	$1.05 \cdot 10^{-17}$	$6.31 \cdot 10^{-4}$

6 Testing the Key Equivalence Hypothesis

Next to the practical security approach, a second important issue for the evaluation of linear cryptanalysis is the key equivalence hypothesis. We consequently decided to investigate the variances of the best linear approximations of a cipher:

$$\underset{\tilde{K}}{\mathbf{var}} \; \max_{\mathbf{a},\mathbf{b}} \; LP(\mathbf{a},\mathbf{b};\tilde{K}) \tag{12}$$

The results of these experiments are summarized in Tables 3 and 4 for various SPNs. As for the previously computed mean values, we observed that these variances decrease with the block size n. A similar observation holds for the variation coefficients (*i.e.* standard deviation over mean). This suggests that the key equivalence hypothesis is reasonable and should not be an issue for SPNs of large (*i.e.* practical) block sizes. We note that it is not necessarily the case for block ciphers with a more "exotic" structure, as illustrated in the case of RC5 and RC6 in [4, 30, 31]. On the other hand, experiments by Selçuk [31] on SP-structured Feistel ciphers seem to match ours.

Table 3: Standard deviation of the best linear probability with a bad diffusion layer.

# rounds	2×4	3×4	2×6	4×4	2×8
1	0	0	0	0	0
2	$5.44 \cdot 10^{-2}$	$5.47 \cdot 10^{-2}$	$5.69 \cdot 10^{-3}$	0	$8.82 \cdot 10^{-4}$
3	$2.39 \cdot 10^{-2}$	$8.23 \cdot 10^{-3}$	$2.54 \cdot 10^{-3}$	$1.15 \cdot 10^{-2}$	$1.25 \cdot 10^{-4}$
4	$1.14 \cdot 10^{-2}$	$7.05 \cdot 10^{-3}$	$6.34 \cdot 10^{-4}$	$7.77 \cdot 10^{-3}$	$4.04 \cdot 10^{-5}$
5	$9.71 \cdot 10^{-3}$	$3.52 \cdot 10^{-3}$	$5.88 \cdot 10^{-4}$	$3.61 \cdot 10^{-3}$	$3.88 \cdot 10^{-5}$
6	$9.65 \cdot 10^{-3}$	$1.51 \cdot 10^{-3}$	$6.06 \cdot 10^{-4}$	$1.55 \cdot 10^{-3}$	$3.87 \cdot 10^{-5}$
7	$9.69 \cdot 10^{-3}$	$6.48 \cdot 10^{-4}$	$6.11 \cdot 10^{-4}$	$7.38 \cdot 10^{-4}$	$3.79 \cdot 10^{-5}$
8	$9.66 \cdot 10^{-3}$	$6.10 \cdot 10^{-4}$	$6.05 \cdot 10^{-4}$	$3.47 \cdot 10^{-4}$	$3.67 \cdot 10^{-5}$

Importantly, these key dependencies highlight another limitation of the provable security approach. Namely, the best linear approximations of a given cipher are not only computationally hard to find, they also only work for one key. It makes them hard to exploit by an actual adversary. This fact is typically used in the decorrelation theory [33]. Otherwise said, these best approximations are generally not relevant to the practical security of ciphers.

Table 4: Standard deviation of the best linear probability with a good diffusion layer.

# rounds	2×4	3×4	2×6	4×4	2×8
1	0	0	0	0	0
2	$8.68 \cdot 10^{-3}$	$3.81 \cdot 10^{-3}$	$1.08 \cdot 10^{-3}$	$3.51 \cdot 10^{-4}$	$9.56 \cdot 10^{-5}$
3	$9.56 \cdot 10^{-3}$	$6.42 \cdot 10^{-4}$	$6.16 \cdot 10^{-4}$	$5.21 \cdot 10^{-5}$	$3.80 \cdot 10^{-5}$
4	$9.64 \cdot 10^{-3}$	$6.15 \cdot 10^{-4}$	$5.81 \cdot 10^{-4}$	$3.88 \cdot 10^{-5}$	$4.09 \cdot 10^{-5}$
5	$9.67 \cdot 10^{-3}$	$6.13 \cdot 10^{-4}$	$6.09 \cdot 10^{-4}$	$3.72 \cdot 10^{-5}$	$3.50 \cdot 10^{-5}$
6	$9.66 \cdot 10^{-3}$	$6.05 \cdot 10^{-4}$	$6.10 \cdot 10^{-4}$	$4.10 \cdot 10^{-5}$	$4.07 \cdot 10^{-5}$
7	$9.64 \cdot 10^{-3}$	$6.10 \cdot 10^{-4}$	$6.14 \cdot 10^{-4}$	$3.50 \cdot 10^{-5}$	$3.92 \cdot 10^{-5}$
8	$9.71 \cdot 10^{-3}$	$6.11 \cdot 10^{-4}$	$6.19 \cdot 10^{-4}$	$3.91 \cdot 10^{-5}$	$3.67 \cdot 10^{-5}$

7 Towards actual design criteria

As previously mentioned, proving the security against linear cryptanalysis would ideally require to relax the practical security approach and to investigate the best linear probability values for any given cipher. Because of computational complexity and key-dependencies, this is hardly achievable in practice. In this section, we aim to take advantage of our computationally tractable block cipher sizes to put forward how theoretical aspects in linear cryptanalysis could possibly lead to practical design criteria. In particular, we aim to determine the number of rounds required for a given cipher to reach its minimum $\mathbf{E}_{\tilde{K}} \max LP$ value and to illustrate the intuitive consequences of reaching it. For these purposes, we computed the mean and variances of the best linear probabilities for different ciphers, over the keys (as previously done) and over the number of rounds:

$$\underset{\tilde{K}}{\mathbf{E}} \max_{\mathbf{a},\mathbf{b}} LP(\mathbf{a},\mathbf{b};\tilde{K}) \tag{13}$$

$$\underset{R}{\mathbf{E}} \max_{\mathbf{a},\mathbf{b}} LP(\mathbf{a},\mathbf{b};\tilde{K}) \tag{14}$$

$$\underset{\tilde{K}}{\mathbf{var}} \max_{\mathbf{a},\mathbf{b}} LP(\mathbf{a},\mathbf{b};\tilde{K}) \tag{15}$$

$$\underset{R}{\mathbf{var}} \max_{\mathbf{a},\mathbf{b}} LP(\mathbf{a},\mathbf{b};\tilde{K}) \tag{16}$$

The variance over the rounds was obtained by implementing our ciphers with R varying between 20 (where it is assumed that the cipher is practically secure) and 10020. The results of these experiments are summarized in our previous tables (for means and variances over the key values) and in the upper part of Table 5 (for means and variances over the number of rounds). They yield the interesting intuition that after a sufficient number of rounds, adding a round and changing the key are statistically undistinguishable, at least from the mean and variance points of view. We consequently derived the following definition:

Definition 5. *A block cipher is in its stationary area with respect to linear approximations if the mean and variance of its average maximum linear probability (over the keys) do not vary in function of the number of rounds R.*

In the lower part of Table 5, we computed the number of rounds for reaching this stationary area, by means of statistical tests for the equality of the mean and variance[3] [5], *i.e.* the number of rounds after which the provable security of a cipher does not evolve anymore. From a designer's point of view, detecting this stationary area is of particular interest since it allows determining precisely the number of rounds that are useful in a block cipher to behave as a "good" n-bit S-box. It would consequently lead to an alternative design criteria for block ciphers to combine, *e.g.* with the wide-trail strategy. But again, from a practical point of view, finding the stationary area is computationally unfeasible for large block sizes. A open question is therefore to determine the number of rounds required to reach this stationary region with tractable heuristics, *e.g.* using the concept of characteristics and/or statistical sampling.

Table 5: Mean and standard deviation of the best linear characteristic for various SPNs over the number of rounds and number of rounds for reaching the "stationary area".

	2×4		3×4		2×6		4×4		2×8	
$\mathbf{E}_R \max(LP)$	$7.64 \cdot 10^{-2}$		$7.44 \cdot 10^{-3}$		$7.43 \cdot 10^{-4}$		$6.30 \cdot 10^{-4}$		$6.30 \cdot 10^{-4}$	
$\mathbf{var}_R \max(LP)$	$9.68 \cdot 10^{-3}$		$6.09 \cdot 10^{-4}$		$6.08 \cdot 10^{-4}$		$3.75 \cdot 10^{-5}$		$3.79 \cdot 10^{-5}$	
Diffusion	bad	good	bad	good	bad	good	bad	good	bad	good
Stationary area	5	4	8	4	6	5	> 8	4	4	3

8 Conclusion

Present strategies to prevent the linear cryptanalysis against block ciphers are based upon important hypotheses that we review in this paper. We first illustrate the large distance between the practical security provided by the best linear characteristic in a cipher and its provable security, determined by the best linear approximation of the cipher. Second, we consider the key equivalence hypothesis and experimentally confirm its validity for reasonably designed ciphers. These results highlight the relevance of a practical approach in block cipher design that does not prevent the existence of good linear approximations but makes them hard to find/exploit. They put forward the interest of better understanding theoretical aspects in linear cryptanalysis and the importance of properly assessing the meaning of design techniques such as the wide-trail strategy.

The distance between theoretical and practical aspects in linear cryptanalysis also motivates the research for powerful tools to exploit the existence of linear approximations within actual ciphers. As a matter of fact, the best known methodology allowing to take advantage of the linear hull effect within a cipher is based on multiple linear approximations. This question therefore relates to the recent work of Biryukov *et al.* [3] in which optimistic bounds were provided for linear attacks using multiple approximations. Experimenting with these attacks to clearly evaluate the actual data complexity of a successful linear cryptanalysis therefore appears as a next important step in the study of block ciphers.

Acknowledgements

The authors would like to thank Michaël Quisquater, for giving us the idea of this work, and for several long and helpful discussions.

Notes

[1] Note that in our target key alternating ciphers using a bitwise XOR key addition, $LCP(\Omega, \tilde{K})$ is independent of the key vector.

[2] A typical illustration of this practical approach is the wide-trail strategy [6] in which the block cipher designers ensure that (1) non-linear components within the cipher (*e.g.* S-boxes) have low linear probabilities and (2) any characteristic involves a high number of active non-linear components.

[3] To compare variances, we used the Levene test, which seemed the most appropriate in our case (unknown non-normal distributions).

References

1. Ross J. Anderson, editor. *Fast Software Encryption, Cambridge Security Workshop, Cambridge, UK, December 9-11, 1993, Proceedings*, volume 809 of *Lecture Notes in Computer Science*. Springer, 1994.
2. Eli Biham, editor. *Advances in Cryptology - EUROCRYPT 2003, International Conference on the Theory and Applications of Cryptographic Techniques, Warsaw, Poland, May 4-8, 2003, Proceedings*, volume 2656 of *Lecture Notes in Computer Science*. Springer, 2003.
3. Alex Biryukov, Christophe De Cannière, and Michaël Quisquater. On Multiple Linear Approximations. In Franklin [11], pages 1–22.
4. J. Borst, B. Preneel, and J. Vandewalle. Linear Cryptanalysis of RC5 and RC6. In Lars R. Knudsen, editor, *Fast Software Encryption, 6th International Workshop, FSE '99, Rome, Italy, March 24-26, 1999*, volume 1636 of *Lecture Notes in Computer Science*, pages 16–30, Berlin, 1999. Springer-Verlag.
5. B. Brown and A.B. Forsythe. Robust Tests for the Equality of Variances. *Journal of American Statistical Association*, 69(346), 1974.
6. Joan Daemen and Vincent Rijmen. The Wide Trail Design Strategy. In Honary [14], pages 222–238.
7. Joan Daemen and Vincent Rijmen. *The Design of Rijndael: AES - The Advanced Encryption Standard*. Springer, 2002.
8. Joan Daemen and Vincent Rijmen. Probability distributions of Correlation and Differentials in Block Ciphers. Cryptology ePrint Archive, Report 2005/212, 2005. `http://eprint.iacr.org/`.
9. Donald W. Davies, editor. *Advances in Cryptology - EUROCRYPT '91, Workshop on the Theory and Application of Cryptographic Techniques, Brighton, UK, April 8-11, 1991, Proceedings*, volume 547 of *Lecture Notes in Computer Science*. Springer, 1991.
10. Yvo Desmedt, editor. *Advances in Cryptology - CRYPTO '94, 14th Annual International Cryptology Conference, Santa Barbara, California, USA, August 21-25, 1994, Proceedings*, volume 839 of *Lecture Notes in Computer Science*. Springer, 1994.

11. Matthew K. Franklin, editor. *Advances in Cryptology - CRYPTO 2004, 24th Annual International CryptologyConference, Santa Barbara, California, USA, August 15-19, 2004, Proceedings*, volume 3152 of *Lecture Notes in Computer Science*. Springer, 2004.
12. Dieter Gollmann, editor. *Fast Software Encryption, Third International Workshop, Cambridge, UK, February 21-23, 1996, Proceedings*, volume 1039 of *Lecture Notes in Computer Science*. Springer, 1996.
13. Carlo Harpes, Gerhard G. Kramer, and James L. Massey. A Generalization of Linear Cryptanalysis and the Applicability of Matsui's Piling-Up Lemma. In L.C. Guillou and J.-J. Quisquater, editors, *Advances in Cryptology - EUROCRYPT '95, Saint-Malo, France, May 21-25, 1995, Proceedings*, pages 24–38, 1995.
14. Bahram Honary, editor. *Cryptography and Coding, 8th IMA International Conference, Cirencester, UK, December 17-19, 2001, Proceedings*, volume 2260 of *Lecture Notes in Computer Science*. Springer, 2001.
15. Pascal Junod. On the Optimality of Linear, Differential, and Sequential Distinguishers. In Biham [2], pages 17–32.
16. Liam Keliher, Henk Meijer, and Stafford E. Tavares. New Method for Upper Bounding the Maximum Average Linear Hull Probability for SPNs. In Pfitzmann [26], pages 420–436.
17. Lars R. Knudsen. Practically Secure Feistel Ciphers. In Anderson [1], pages 211–221.
18. Lars R. Knudsen, editor. *Fast Software Encryption, 6th International Workshop, FSE '99, Rome, Italy, March 24-26, 1999, Proceedings*, volume 1636 of *Lecture Notes in Computer Science*. Springer, 1999.
19. Lars R. Knudsen and John Erik Mathiassen. On the Role of Key Schedules in Attacks on Iterated Ciphers. In Samarati et al. [28], pages 322–334.
20. Lars R. Knudsen and Vincent Rijmen. On the Decorrelated Fast Cipher (DFC) and Its Theory. In Lars R. Knudsen, editor, *Fast Software Encryption, 6th International Workshop, FSE '99, Rome, Italy, March 24-26, 1999*, volume 1636 of *Lecture Notes in Computer Science*, pages 81–94. Springer-Verlag, 1999.
21. Xuejia Lai and James L. Massey. Markov Ciphers and Differentail Cryptoanalysis. In Davies [9], pages 17–38.
22. Mitsuru Matsui. The First Experimental Cryptanalysis of the Data Encryption Standard. In Desmedt [10], pages 1–11.
23. Mitsuru Matsui. New Structure of Block Ciphers with Provable Security against Differential and Linear Cryptanalysis. In Gollmann [12], pages 205–218.
24. Kaisa Nyberg. Linear Approximation of Block Ciphers. In Santis [29], pages 439–444.
25. Kaisa Nyberg and Lars R. Knudsen. Provable Security Against a Differential Attack. *J. Cryptology*, 8(1):27–37, 1995.
26. Birgit Pfitzmann, editor. *Advances in Cryptology - EUROCRYPT 2001, International Conference on the Theory and Application of Cryptographic Techniques, Innsbruck, Austria, May 6-10, 2001, Proceedings*, volume 2045 of *Lecture Notes in Computer Science*. Springer, 2001.
27. V. Rijmen. *Cryptanalysis and Design of Iterated Block Ciphers.* PhD thesis, KULeuven, October 1997.
28. Pierangela Samarati, Peter Y. A. Ryan, Dieter Gollmann, and Refik Molva, editors. *Computer Security - ESORICS 2004, 9th European Symposium on Research Computer Security, Sophia Antipolis, France, September 13-15, 2004, Proceedings*, volume 3193 of *Lecture Notes in Computer Science*. Springer, 2004.

29. Alfredo De Santis, editor. *Advances in Cryptology - EUROCRYPT '94, Workshop on the Theory and Application of Cryptographic Techniques, Perugia, Italy, May 9-12, 1994, Proceedings*, volume 950 of *Lecture Notes in Computer Science*. Springer, 1995.
30. Ali Aydin Selçuk. New Results in Linear Cryptanalysis of RC5. In Serge Vaudenay, editor, *Fast Software Encryption, 5th International Workshop, FSE '98, Paris, France, March 23-25, 1998*, volume 1372 of *Lecture Notes in Computer Science*, pages 1–16, Berlin, 1998. Springer-Verlag.
31. Ali Aydin Selçuk. On Bias Estimation in Linear Cryptanalysis. In B.K. Roy and E. Okamoto, editors, *Progress in Cryptology - INDOCRYPT 2000, First International Conference in Cryptology in India, Calcutta, India, December 10-13, 2000*, volume 1977 of *Lecture Notes in Computer Science*, pages 52–66, Berlin, 2000. Springer-Verlag.
32. Serge Vaudenay. On the Security of CS-Cipher. In Knudsen [18], pages 260–274.
33. Serge Vaudenay. Decorrelation: A Theory for Block Cipher Security. *J. Cryptology*, 16(4):249–286, 2003.
34. David Wagner. The Boomerang Attack. In Lars R. Knudsen, editor, *Fast Software Encryption, 6th International Workshop, FSE '99, Rome, Italy, March 24-26, 1999*, volume 1636 of *Lecture Notes in Computer Science*, pages 156–170. Springer-Verlag, 1999.

A Detailed Specification of the Cipher Component Used

A.1 The Optimal Diffusion Layers

The optimal diffusion layers we use are based on MDS linear codes constructed from a Vandermonde Matrix (see for example [27] for details). The output of each $n \times n$ S-box is considered as an element of $\mathrm{GF}(2^n)$. Applying the linear layer is equivalent to perform a matrix multiplication with the Vandermonde matrix. Arithmetics is performed in $\mathrm{GF}(2^n)$. We construct the finite field $\mathrm{GF}(2^n)$ as $\mathrm{GF}(2)[\alpha]/(p(\alpha))$, where $\mathrm{GF}(2)[\alpha]$ is the ring of polynomials in one variable α with coefficients in $\mathrm{GF}(2)$, and p is a primitive polynomial of degree n, which is:

- $\alpha^4 \oplus \alpha \oplus 1$ for $n = 4$.
- $\alpha^6 \oplus \alpha \oplus 1$ for $n = 6$.
- $\alpha^8 \oplus \alpha^4 \oplus \alpha^3 \oplus \alpha^2 \oplus 1$ for $n = 8$.

The Vandermonde matrices are:

- If 2 S-boxes are used in each round:

$$\begin{pmatrix} 1 & \alpha \\ 1 & \alpha^2 \end{pmatrix}$$

- If 3 S-boxes are used in each round:

$$\begin{pmatrix} 1 & \alpha & \alpha^2 \\ 1 & \alpha^2 & \alpha^4 \\ 1 & \alpha^3 & \alpha^6 \end{pmatrix}$$

– If 4 S-boxes are used in each round:

$$\begin{pmatrix} 1 & \alpha & \alpha^2 & \alpha^3 \\ 1 & \alpha^2 & \alpha^4 & \alpha^6 \\ 1 & \alpha^3 & \alpha^6 & \alpha^9 \\ 1 & \alpha^4 & \alpha^8 & \alpha^{12} \end{pmatrix}$$

A.2 The "Wire Crossing" Diffusion Layers

We also used poor diffusion layers in our experiments, made of simple bit permutations. These layers are (numbers refer to bit positions, from left to right):

Poor diffusion layer: Two 4 × 4 S-boxes

Input	0	1	2	3	4	5	6	7
Output	0	1	4	5	2	3	6	7

Poor diffusion layer: Two 6 × 6 S-boxes

Input	0	1	2	3	4	5	6	7	8	9	a	b
Output	0	1	2	6	7	8	3	4	5	9	a	b

Poor diffusion layer: Three 4 × 4 S-boxes

Input	0	1	2	3	4	5	6	7	8	9	a	b
Output	0	1	4	8	2	5	6	9	3	7	a	b

Poor diffusion layer: Two 8 × 8 S-boxes

Input	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
Output	0	1	2	3	8	9	a	b	4	5	6	7	c	d	e	f

Poor diffusion layer: Four 4 × 4 S-boxes

Input	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
Output	0	4	8	c	1	5	9	d	2	6	a	e	3	7	b	f

A.3 The S-boxes

We used three different S-boxes, of respective size 4 × 4, 6 × 6, 8 × 8.

The 4 × 4 S-box.

Input	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
Output	b	9	1	d	8	f	0	6	4	c	2	3	e	5	a	7

The 6 × 6 S-box.

	00	01	02	03	04	05	06	07	08	09	0a	0b	0c	0d	0e	0f
00	24	20	21	35	3e	37	d	26	9	33	0	28	27	a	23	31
10	14	3f	13	2e	30	2a	3b	12	34	16	17	18	8	1b	e	1f
20	3c	2	b	32	2d	3	7	1e	2f	1c	10	1d	11	f	25	c
30	6	1	39	15	4	3d	3a	19	1a	36	2b	5	38	2c	29	22

The 8 × 8 S-box.

	00	01	02	03	04	05	06	07	08	09	0a	0b	0c	0d	0e	0f
00	5c	11	85	da	95	c4	f8	49	b2	db	94	c6	1c	c0	22	9
10	b8	24	7c	3a	ad	13	89	7f	c2	84	2e	74	e0	83	c	e
20	eb	ec	6c	ac	57	a	79	3e	41	ce	1a	68	69	16	f7	8c
30	b3	6b	7e	a4	ae	1f	77	ff	5a	65	25	bb	fa	b6	c1	6d
40	8f	d	56	2a	3b	29	6e	ee	19	88	15	d2	9a	98	dd	72
50	1e	9d	32	a0	fb	fc	80	a8	54	ba	51	f6	20	f9	ca	d5
60	d8	f5	78	5d	e9	1	92	5	cb	bf	be	40	2f	f2	a7	df
70	63	48	70	9f	82	b1	8a	35	52	a2	e6	f	76	c3	bc	b7
80	17	d3	af	ab	53	75	3d	de	ed	1b	9e	e8	c9	e2	86	0
90	91	37	fe	64	5f	59	cd	e3	39	a1	7	61	8e	90	7b	23
a0	c5	a5	e7	38	71	8	4b	7d	1d	67	8b	e5	4c	f1	44	a3
b0	f4	55	87	62	d4	46	a9	4a	97	c7	e4	d1	12	81	b4	2b
c0	42	3f	9c	50	4d	aa	6	3	31	58	d0	14	21	b9	d7	6a
d0	30	4	a6	f3	9b	cf	93	96	2	3c	4f	28	4e	27	2c	b0
e0	47	d6	60	ea	5e	26	10	e1	45	cc	2d	7a	6f	33	66	34
f0	f0	8d	fd	c8	5b	36	bd	ef	b5	43	dc	d9	b	73	99	18